Complex Disorders: Populations and Pedigrees

Dr Carole Sargent
Department of Pathology

October 17th 2015

- This lecture – look at non-parametric analysis and the principles behind this approach
- Next lecture- look in more depth at the development of SNP chips and how they can be applied to find the contributing loci.

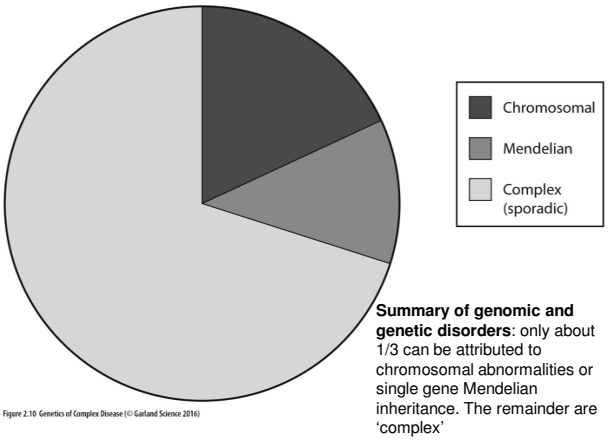
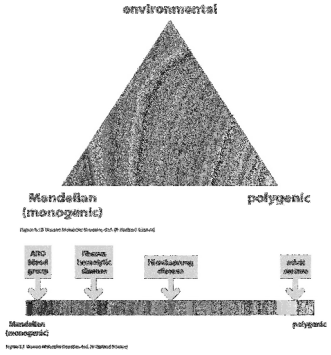


Figure 2.10 Genetics of Complex Disease (© Garland Science 2016)

Summary of genomic and genetic disorders: only about 1/3 can be attributed to chromosomal abnormalities or single gene Mendelian inheritance. The remainder are 'complex'

What is a complex disorder?

- Complex/ multifactorial/ polygenic disorders
- May also have environmental influence on outcome
- ABO: single locus
- Rhesus: interaction between mother and baby
- Height multiple genes BUT environment (e.g. nutrition) also contributes.



Distribution of genotypes/ phenotypes

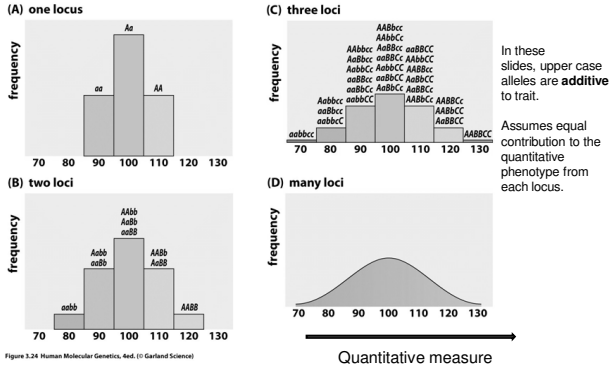
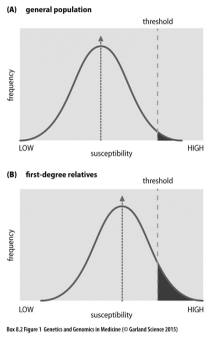


Figure 3.24 Human Molecular Genetics, 4ed. (© Garland Science)

In these slides, upper case alleles are **additive** to trait.

Assumes equal contribution to the quantitative phenotype from each locus.

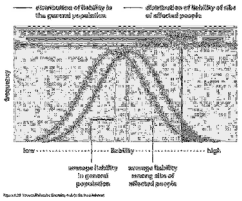
Complex phenotype: population based studies v pedigrees



Box 8.2 Figure 1 Genetics and Genomics in Medicine (© Garland Science 2015)

Many complex traits are not quantitative. They are discontinuous (threshold) traits.

This means that the individuals in a population or a pedigree who develop the phenotype must carry a sufficient burden of 'risk' alleles to exceed some threshold at the high end of the susceptibility (liability) distribution.



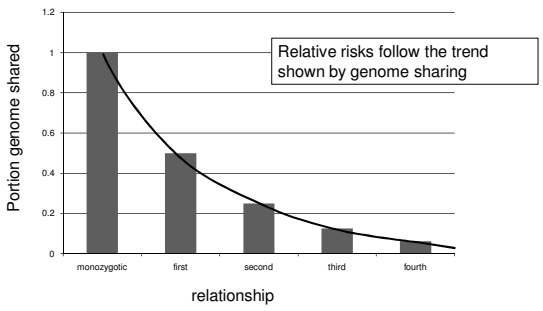
Parametric Analysis: recap

- Good for genetic disorders where variation (mutation) at a single genetic locus contributes ~100% genetic contribution with ~100% penetrance and ~100% expression.
  - *Penetrance= number of individuals in the pedigree that show the phenotype*
  - *Expression= phenotypic variance (ideally no variance)*
- Need to have mode of inheritance
- Need to know approximate frequency of disorder in population
- Can use pedigrees to analyse with LOD scores
- What happens if we don't know the number of loci, penetrance, or other variables to fit to the model?

Mapping genes for complex disorders

- Variance (V): defines the phenotypic range, and includes both genetic and environmental contributions
- Heritability (h^2): defines the proportion of the phenotype that can be attributed to genetics
- We don't know how many genes are involved
- We don't know how much each locus contributes
- We can prove there is a genetic input

Gene mapping Model-free (nonparametric)



DISEASE	RELATIVE RISK FOR SIB (λ_s)
Alzheimer disease (late-onset)	4
Autism spectrum disorder	6.5
Breast cancer, female	2
Crohn's disease	25
Multiple sclerosis	20
Schizophrenia	9
Type 1 diabetes	15
Type 2 diabetes	3

DISEASE	CONCORDANCE (%)	
	In MZ twins	In DZ twins
Type 1 diabetes	42.9	7.4
Type 2 diabetes	34	16
Multiple sclerosis	25.3	5.4
Crohn's disease	37	10
Ulcerative colitis	7	3
Alzheimer disease	32.2	8.7
Parkinson disease	15.5	11.1
Schizophrenia	40.8	5.3

Model-free analysis

- **Linkage analysis – affected sibs or extended pedigrees**
- **Homozygosity mapping (pedigrees)**
- **Association mapping (population based)**
- **Transmission Disequilibrium Test (TDT) (family based verification of association studies)**

Model-free linkage analysis

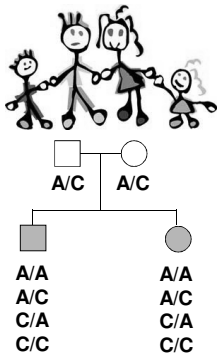
- **Do affected relatives share a chromosomal segment in common more often than would be expected on the basis of their relationship in the pedigree? (shared segment analysis)**
- **Do <u>not</u> need to specify**
  - **-mode of inheritance**
  - **-number of loci**
  - **- gene frequencies**
  - **- penetrance**

Model-free vs model-based linkage analysis

- Can use smaller family clusters (e.g. affected sib-pairs) rather than larger pedigrees
- More robust to errors. Model-based analysis will be sensitive to errors in the model.
- <u>But</u> less powerful than model-based analysis. Need more individuals to achieve statistical significance.

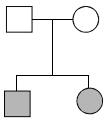
(*Power of the experiment is the chance of finding the trait-causing genes e.g. 80-90% power is good*)

Affected sib-pair analysis



		Alleles in common		
sib 1	sib 2	0	1	2
A/A	A/A			+
A/A	A/C		+	
A/A	C/A		+	
A/A	C/C	+		
A/C	A/C			+
A/C	A/A		+	
A/C	C/C		+	
A/C	C/A	+		
C/A	C/A			+
C/A	C/C		+	
C/A	A/A		+	
C/A	A/C	+		
C/C	C/C			+
C/C	C/A		+	
C/C	A/C		+	
C/C	A/A	+		
TOTALS		4	8	4

Affected sib-pair analysis

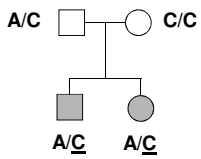


Alleles in common
0 1 2

Expected ratio (no linkage)
1 : 2 : 1

Linkage to a susceptibility locus increases the probability of having alleles in common and frequency of shared alleles exceeds expectation

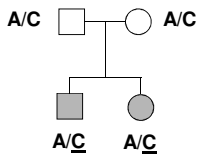
Affected sib-pair analysis



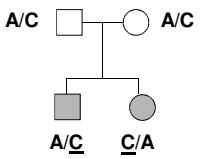
Identical-by-descent (IBD)

- **If the allele associated with the phenotype is inherited from dad, we can be certain that the affected sibs share a segment of chromosome carrying 'A'**
- **If the phenotype is inherited from mum, marker 'C' is not informative, as it could come from either of her chromosomes**

Affected sib-pair analysis



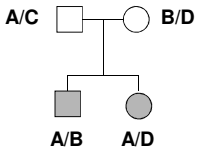
Identical-by-descent (IBD)



Identical-by-state (IBS)

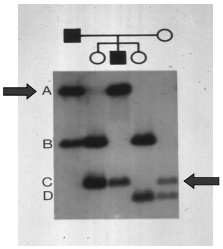
- **Genotypes are the same, but the origins of the alleles are different**
- **Single SNPs do not give sufficient information**

Affected sib-pair analysis

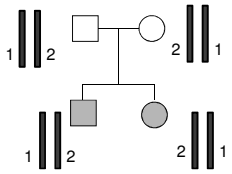


Identical-by-descent (IBD)

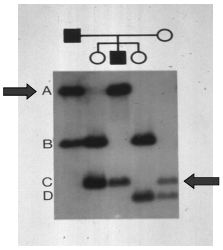
- **Using markers with multiple alleles (microsatellites or haplotypes based on multiple SNPS) we can be more confident of IBD**
- **Genotyping the parents helps, as it established the phase of inheritance**



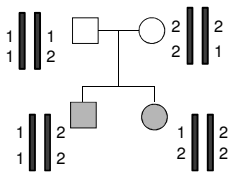
MS: 4 alleles, easy to follow through the pedigree. Allele **A** is co-inherited with the phenotype



One SNP, only 2 alleles.
Cannot determine IBD v IBS



MS: 4 alleles, easy to follow through the pedigree. Allele **A** is co-inherited with the phenotype



Two SNPs, four possible haplotypes (11,12, 21, 22).
The more loci included in the haplotype, the more 'alleles' are generated, making it easier to determine IBD v IBS.

Affected sib-pair analysis: Summary 1

Null hypothesis
H0: $(p_0, p_1, p_2) = (¼, ½, ¼)$
H1: $(p_0, p_1, p_2) \neq (¼, ½, ¼)$

ASP tests can be broadly classified into:

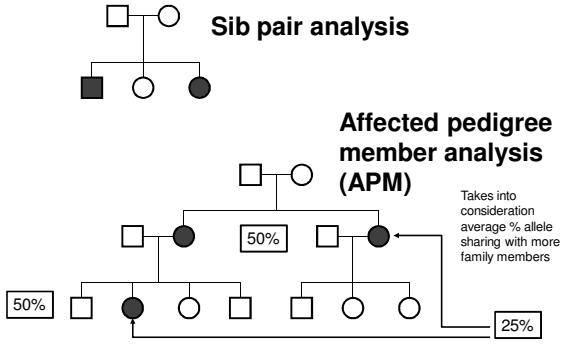
(1) score tests based on counting observed number of ASPs sharing 0,1 or 2 alleles, including chi-squared goodness of fit – assume IBD status can be unequivocally resolved

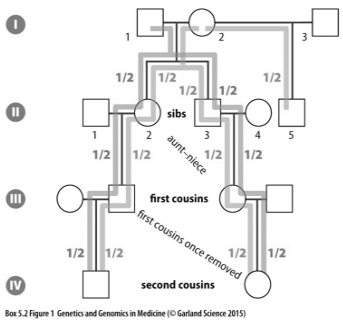
(2) tests based on likelihood statistics

Affected sib-pair analysis: Summary 2

- **Want to estimate IBD sharing**
- **Need highly polymorphic markers (usually microsatellites)**
- **Or use SNPs to give haplotypes**
- ***Likelihood statistics can be used when IBD status cannot be fully resolved***

Extending model-free linkage analysis





Box 5.2 Figure 1 Genetics and Genomics in Medicine (© Garland Science 2015)

proportion of genes in common

parent–child: 1/2

grandparent–grandchild: 1/2 x 1/2 = 1/4

sibs: 1/2 x 1/2 + 1/2 x 1/2
= 1/4 + 1/4 = 1/2

half sibs (for example II-3 and II-5): 1/2 x 1/2 = 1/4

uncle/aunt–nephew/niece: $(1/2)^3 + (1/2)^3$
= 1/8 + 1/8 = 1/4

first cousins: $(1/2)^4 + (1/2)^4$
= 1/16 + 1/16 = 1/8

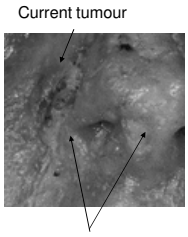
first cousins once removed: $(1/2)^5 + (1/2)^5$
= 1/32 + 1/32 = 1/16

second cousins: $(1/2)^6 + (1/2)^6$
= 1/64 + 1/64 = 1/32

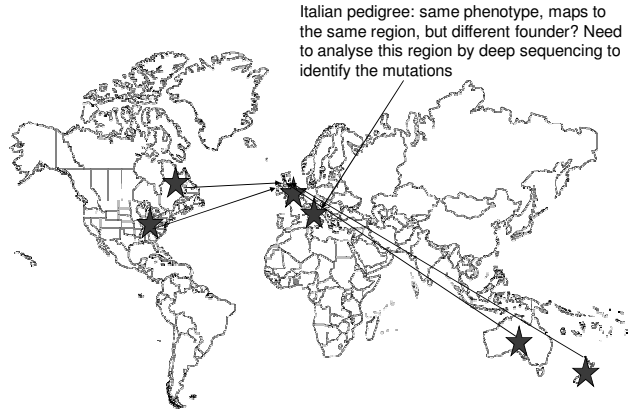
Understanding the proportion of sharing means the null hypothesis can be applied to any pairwise comparisons in an APM.

Case Study: MSSE (multiple squamous self healing epithelioma, or Ferguson-Smith disease) – single major locus

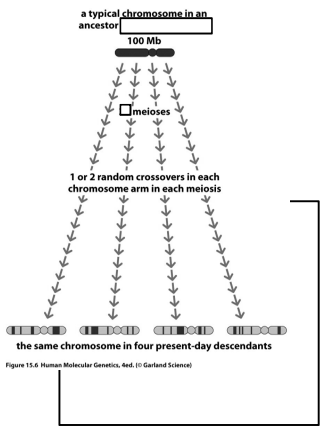
- First described in 1934
- Tumours on exposed skin (UV/radiation → increased incidence: environmental impact)
- Self resolve but leave pitted scars which can disfigure
- Variable age of onset (6-60+)
- Dominant
- Not fully penetrant
- Can't be sure who is true 'unaffected' member of pedigree
- Genomic interval first defined by in 1993 based on LOD scores
- MS sharing defined limits of region of chromosome 9

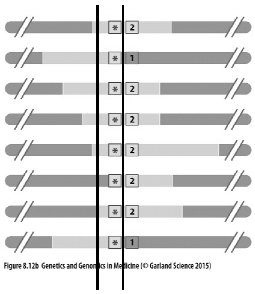


Adapted from S. J. Robertson et al, Clinical and Experimental DermatologyClinical dermatology Vol. 35: e100-e102



Many cases worldwide belong to an extended pedigree with a founder from the 1700s who was a member of the Ayrshire coal-mining community. Affected individuals share a haplotype of markers clustered on chromosome 9q22-q31.



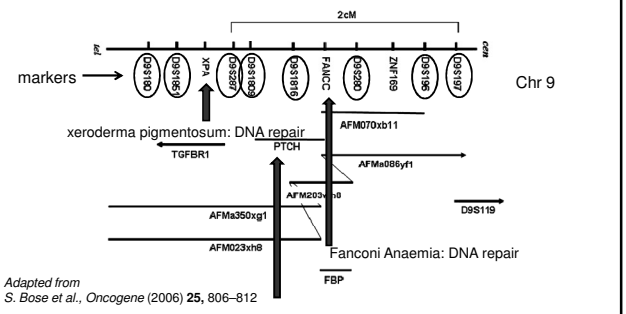


Recombination reduces the size of the shared fragment in each generation.

Each individual inherits slightly differently sized blocks of the original chromosomal region

Region of common overlap defines region to investigate.

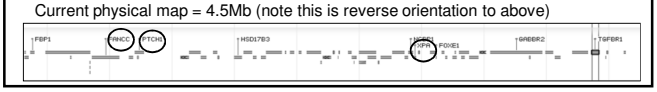
Interval selected using genetic data



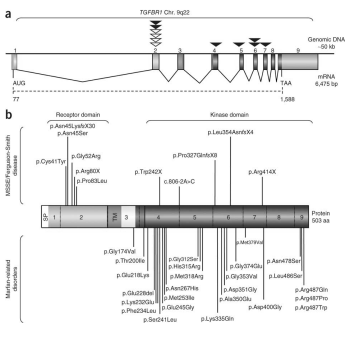
Adapted from
S. Bose et al., Oncogene (2006) **25,** 806–812

PATCHED 1: known to be involved in basal cell carcinoma

Current physical map = 4.5Mb (note this is reverse orientation to above)



Role of TGFBR1



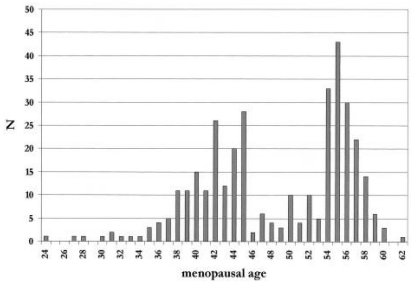
From Nature Genetics 43:365-371

- Mutations identified as the changes were not previously observed in the SNP database.
- All families had mutations in *TGFBR1* that result in classic TS like haplo-insufficiency at this locus. Tumours arise when the normal allele is mutated
- Other mutations in this gene cause Marfan syndrome related disorders of cardiovascular system, neurocognitive development and craniofacial-skeletal development. These are all autosomal dominant mutations. *NOTE all aa substitutions*

Model-free linkage analysis for quantitative traits

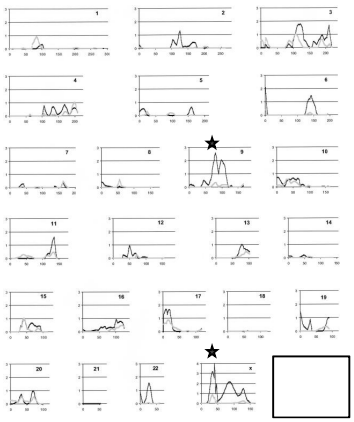
- Quantitative traits show population variance, but can be given a defined value
- Determining the number of genes, and position of contributing genes is based on IBD sharing between relatives

Age of Menopause



- Heritability h^2 is 31-85%
- families with age of onset <45 (early) or >54 (late)
- scan genome with MS markers

Asselt *et al.* (2004) Am J Hum Genet 74: 444-453

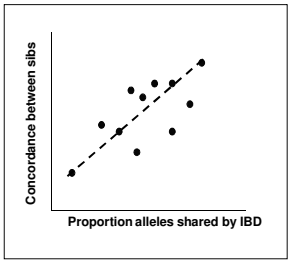


Data from MS scan analysed to generate LOD scores using standard methods and regression analysis to look for the strongest chromosomal contributions.

Asselt *et al.* (2004) Am J Hum Genet 74: 444-453

Model-free linkage analysis for quantitative traits

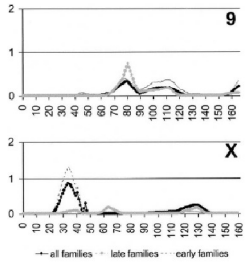
- **Sib-pair regression (Haseman-Elston)**



Regression analysis shows that the more similar the sib-pairs are, the more alleles are shared between them.

This is expected based on predictions for a quantitative trait.

Asselt *et al.* (2004) Am J Hum Genet 74: 444-453



Both chromosomal regions remain positive with the regression analysis but with reduced LOD scores.

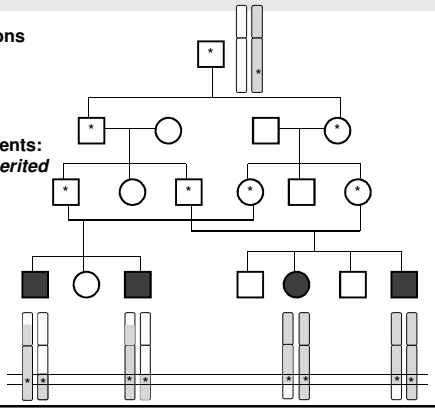
Both chromosomal regions encode proteins of the BCL2 family, which are involved in the control of apoptosis.

This may relate to the potential number of mature oocytes at birth.

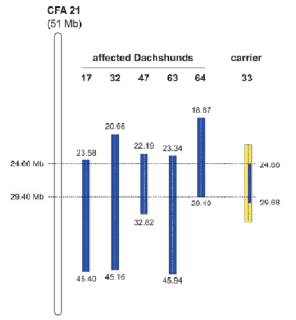
Asselt *et al.* (2004) Am J Hum Genet 74: 444-453

Homozygosity mapping

- **recessive conditions**
- **inbred pedigrees**
- **search for shared homozygous segments:** ***both alleles are inherited from a common ancestor***



Homozygosity mapping

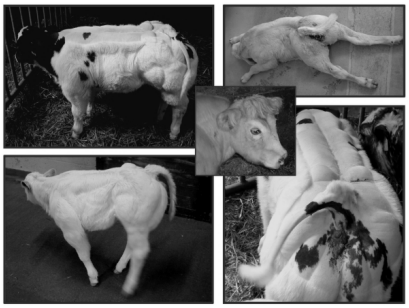


- **Osteogenesis Imperfecta**
- **certain pedigree lineages of Dachshund dogs**
- **Fragile teeth and bones**
- **MS mapping provides evidence for a single locus**
- **Region identified is large**
- **Add SNP data to the study**
- **SERPINH1 gene (involved with collagen folding)**

Drogemuller *et al.* (2009) PLoS Genetics 5: e1000579

Regions of homozygosity defined in each animal, delimited by a common haplotype seen in heterozygous carrier. Study used limited panel of SNPs

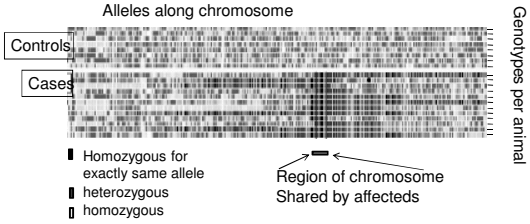
Homozygosity mapping: crooked tail syndrome in Belgian Blue cattle



Clinical spectrum exhibited by CTS cases.
Crooked tail, growth retardation, stocky head, extreme muscular hypertrophy, spastic paresis of the hind limbs, straight hock, scoliosis.

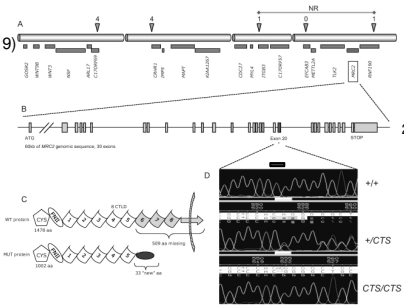
See PLoS Genet. 2009 September; 5(9): e1000666.

Figure adapted from example of cattle recessive disease mapping in Nature Genetics 2008, 40:449



Identify area of chromosome that is homozygous in cases but not controls: must have been inherited from original mutated chromosome. This form of homozygosity mapping is also called ***autozygosity mapping.*** The smallest common interval is the target for further investigation.

(Cattle chr 19)



Region defined

Potential candidates: sequence all 7 in a case and control

2bp deletion in *MRC2*

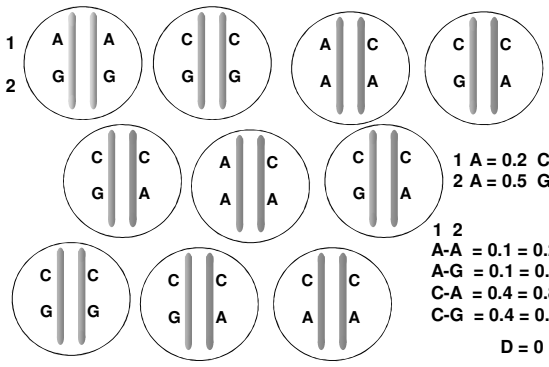
The encoded MRC2 protein plays a role in extracellular matrix remodelling by mediating the internalization and lysosomal degradation of collagen ligands. Collagen is a component of the skeleton, and this explains the phenotype.

Heterozygous carriers have more muscle and have been selected as breeding sires over their homozygous wild type sibs. The frequency of the mutant allele has increased in the population as a result of this selection: ***SELECTIVE SWEEP.***

See PLoS Genet. 2009 September; 5(9): e1000666.

Association: population based studies

Population studies rely on Linkage disequilibrium

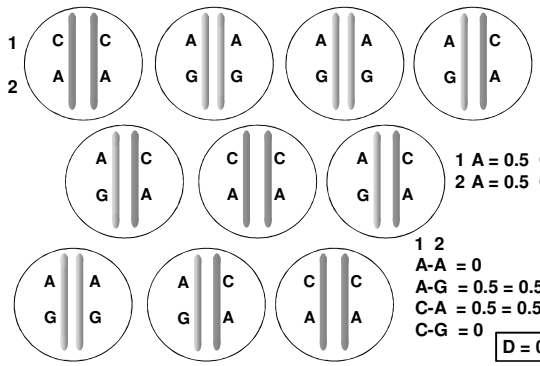


1 A = 0.2 C = 0.8
2 A = 0.5 G = 0.5

1 2
A-A = 0.1 = 0.2 x 0.5
A-G = 0.1 = 0.2 x 0.5
C-A = 0.4 = 0.8 x 0.5
C-G = 0.4 = 0.8 x 0.5

D = 0

Linkage disequilibrium



1 A = 0.5 C = 0.5
2 A = 0.5 G = 0.5

1 2
A-A = 0
A-G = 0.5 = 0.5 x 0.5 + D
C-A = 0.5 = 0.5 x 0.5 + D
C-G = 0

D = 0.25

Linkage disequilibrium

1 2
A-A = 0
A-G = 0.5 = 0.5 x 0.5 + D
C-A = 0.5 = 0.5 x 0.5 + D
C-G = 0

D = 0.25

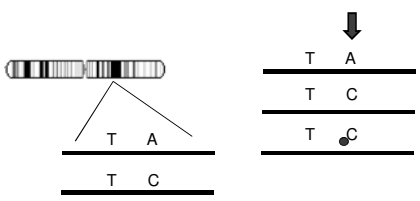
i.e, D = (observed frequency of the haplotype at loci 1&2) –
(expected frequency of the haplotype at loci 1&2)

Where (expected frequency)= (frequency of allele at locus 1) x
(frequency of allele at locus 2)

Assuming independent assortment of alleles

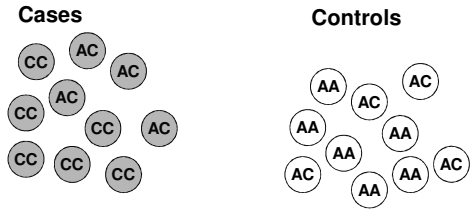
If D ≠ 0, then there must be linkage disequilibrium

Principles of Association analysis



- Simple case: two haplotypes, TA and TC at two adjacent loci
- Polymorphism contributing to trait lies near C on TC background
- Separate into cases and controls.
- Genotype at locus A/C

Association



	Cases	Controls
A	0.20	0.80
C	0.80	0.20

Association mapping (GWAS)

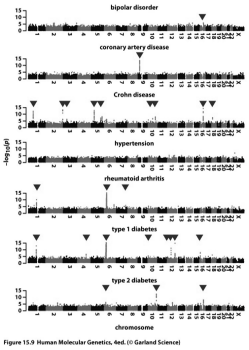
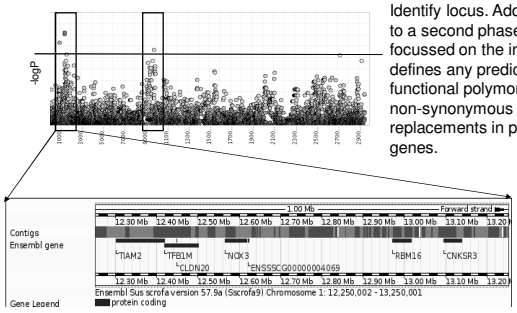


Figure 15.9 Human Molecular Genetics, 4ed. (© Garland Science)

Summary for the major complex disorders investigated for the Wellcome Trust Case Control COnsortium

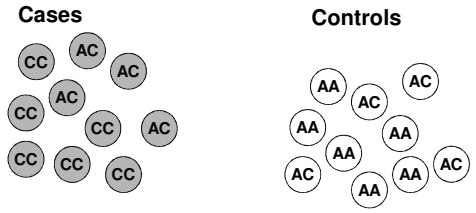
From association data to genes



Identify locus. Add more SNPs to a second phase of analysis focussed on the interval: this defines any predicted functional polymorphisms, e.g. non-synonymous aa replacements in protein coding genes.

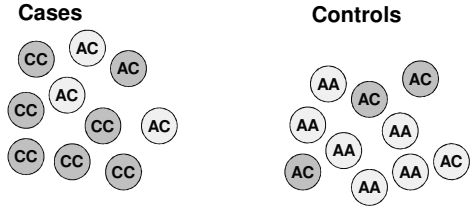
Note: use the –logP value, as we are looking at very small p values in the original data. Transforming on the –log scale makes the significant results more obvious!

Spurious association



	Cases	Controls
A	0.20	0.80
C	0.80	0.20

Spurious association

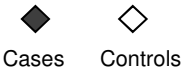


	Chinese	European
A	0.20	0.80
C	0.80	0.20

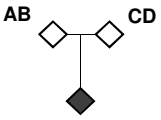
Spurious association

- **Disease susceptibility shows ethnic differences**
- **Polymorphisms show ethnic differences**
- **This can lead to spurious or false associations being detected if samples from different populations are analysed together**

Transmission Disequilibrium Test (TDT)



Case-control study: is allele A more frequent in cases than controls?



Cases + parents

TDT: when a parent has allele A and is heterozygous, is allele A transmitted to the affected offspring more frequently than the expected 50% of times?

TDT avoids spurious association due to population stratification

Linkage vs association mapping

- **Linkage detectable over large genetic distances, typically 10-20 cM with large sample and many informative meioses**
- **Allelic association has to persist over many generations, so only detectable over small genetic distances of the order of 1 cM (humans)**
- **Linkage and association complement each other**
- **New methods of combined linkage and LD analysis enable the two approaches to be combined in a single analysis**
  - **Especially useful with multiple small pedigrees with >1 affected sib**

Linkage and association pros and cons

- **Linkage studies**
  - **Good power for genes of large to medium effect**
  - **Need extremely large samples to detect weak effects**
  - **Computationally difficult with large number of genotypes**
- **Association studies**
  - **More powerful than linkage for small gene effects**
  - **Suitable for high throughput genotyping**
  - **Need to beware of spurious associations**

Reading

Human Molecular Genetics 3 (edition 4 is now available!)
Strachan & Read, Garland Science 2003

Chapter 13 Genetic mapping of Mendelian characters

Chapter 15 Mapping and Identifying Genes Conferring Susceptibility to Complex Diseases

Altshuler, D., Daly, M.J. and Lander, E.S. (2008) Genetic mapping in human disease **Science** *322*: 881-888

References

Asselt *et al.*, **Linkage analysis for age of menopause**, Am J Hum Genet, 2004, Vol. 74, 444 - 453
http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1182258/?tool=pubmed

Drogemuller *et al.*, **A missense mutation in the SERPINH1 gene in Dachshunds with osteogenesis imperfecta**, PLoS Genetics 2009 Vol. 5, e1000579
http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2708911/?tool=pubmed

Weetman *et al.*, **Association mapping of insecticide resistance in wild *Anopheles Gambiae***, PLoS One 2010 Vol. 5, e13140
http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2956759/?tool=pubmed